

---

# Methods

---

## Measuring Hospital Efficiency: A Comparison of Two Approaches

*Thomas N. Chirikos and Alan M. Sear*

---

**Objective.** To compare the results of scoring hospital efficiency by means of two new types of *frontier* models, Data Envelopment Analysis (DEA) and stochastic frontier regression (SFR).

**Study Setting.** Financial records of Florida acute care hospitals in continuous operation over the period 1982–1993.

**Study Design.** Comparable DEA and SFR models are specified, and these models are then estimated to obtain the efficiency indexes yielded by each. The empirical results are subsequently examined to ascertain the extent to which they serve the needs of hospital policymakers.

**Data Collection.** A longitudinal or panel data set is assembled, and a common set of output, input, and cost indicators is constructed to support the estimation of comparable DEA and SFR models.

**Principal Findings.** DEA and SFR models yield convergent evidence about hospital efficiency at the *industry* level, but divergent portraits of the *individual* characteristics of the most and least efficient facilities.

**Conclusions.** Hospital policymakers should not be indifferent to the choice of the *frontier* model used to score efficiency relationships. They may be well advised to wait until additional research clarifies reasons why DEA and SFR models yield divergent results before they introduce these methods into the policy process.

**Key Words.** Hospital efficiency, Data Envelopment Analysis, stochastic frontier regression, hospital cost containment

---

Until recently, the efficiency of hospitals has been measured by estimating cost or production functions by means of ordinary regression methods. Cost studies, for instance, typically regressed operating expenses against various measures of hospital “output,” input prices, and control variables such as case mix of the patient population to draw inferences about efficiency differentials across hospitals (Cowing, Holtmann, and Powers 1983). Although these regression studies produced many useful insights, they were unavoidably subject to the limitation that the estimated equation represented the *average* as opposed to the *best-practice* cost-output relationship. The error term

in such regressions has a mean of zero, so deviations from the estimated "line" (hyperplane) are not only as likely to raise costs as reduce them, but are assumed to be entirely attributable to chance factors. It is intuitively clear that some observations below the regression "line" in cost equations are systematically more efficient in the sense that, for a given set of factor prices and patient characteristics, they produce more output per unit of input. However, ordinary regression methods cannot distinguish between these systematic variations and those truly due to statistical noise. In an industry where inefficiencies are thought to be widespread, this methodological gap seriously limits the use of such statistical inferences for policy purposes.

Recent years have witnessed technical developments in the field of management science and econometrics that hold out the promise of enabling analysts to identify *best-practice* output-input (cost) relationships as well as to gauge how much efficiency levels of given decision-making units or providers deviate from these *frontier* values (Bauer 1990). One of these developments is Data Envelopment Analysis (DEA), a nonparametric programming technique that pieces together an efficiency frontier by maximizing (*a seriatim*) the weighted output/input (cost) ratio of each provider, subject to the condition that this ratio can equal, but never exceed, unity for any other provider in the data set (Charnes, Cooper, and Rhodes 1978). DEA then yields several measures of the relative distance of any provider's efficiency ratio from the piecewise linear frontier, the most common being the proportional reduction in input or cost levels that could be achieved were the provider delivering services in the most efficient manner possible. Another was the development of stochastic frontier regression (SFR) methods. Unlike its classical OLS counterpart, SFR models the error term in two parts, one reflecting systematic deviations from a frontier (cost or output) level and the other from more conventional statistical noise (Aigner, Knox Lovell, and Schmidt 1977). SFR uses this composable error, as it is called, to estimate the overall efficiency

---

With the standard disclaimers, financial support from HCFA Cooperative Agreement No. 17-C-90285/4-01 is gratefully acknowledged. Views expressed in the article are those of the authors alone and do not necessarily reflect the views or policies of the Health Care Financing Administration.

Address correspondence to Thomas N. Chirikos, Professor and Member-in-Residence, H. Lee Moffitt Cancer Center and Research Institute, Cancer Control, Moffitt Research Center, University of South Florida, 12902 Magnolia Drive, Tampa FL 33612-9497. Alan M. Sear, Ph.D. is an Associate Professor, Dept. of Health Policy and Management, University of South Florida. This article, submitted to *Health Services Research* on November 26, 1996, was revised and accepted for publication on January 15, 1999.

level across any sample of providers and then, in what may be characterized as a second step, computes efficiency deviations of each sample observation from the industry frontier by taking the expected value of the disturbance of each observation, conditional on the estimated parameters of the underlying distribution of the composable error (Jondrow et al. 1982; Greene 1992). Like DEA, these SFR provider-specific efficiency values are also cast as the proportional difference between the costs (output) of any provider and the frontier level of costs (output).

Not surprisingly, the efficiency vectors yielded by DEA and SFR techniques find ready uses by policymakers, nowhere more so than in the hospital sector. Indicators of the relative efficiency of hospitals are needed to gauge whether hospital cost-containment efforts are succeeding; they are also needed to evaluate the effect of more extensive managed care arrangements in local healthcare markets and to prepare "report cards" and other quality assessments of hospital service delivery. Such indicators may also have a prescriptive role to play in establishing criteria for selective contracting purposes and in pegging reimbursement levels in hospital rate-setting programs; see Batavia et al. (1993), Hadley and Zuckerman (1994), and Newhouse (1994) for differing views on these potential uses. Yet, whether frontier methods actually live up to their policy promise is untested at the moment. The recent literature includes a number of DEA hospital applications but only one published SFR study.<sup>1</sup> The knowledge base must be expanded before we can judge if, and how well, frontier methods serve the needs of hospital policymakers.

New research should be directed not only on additional DEA- and SFR-specific studies, but also on comparative analyses of the results yielded by each type of model. Comparative studies are needed because DEA and SFR appear to be treated in some parts of the literature as sufficiently close substitutes to afford decision makers a choice and/or a means of cross-validating results, while in other parts they are treated as complementary techniques yielding triangulated results (cf. Batavia et al. 1993; Kooreman 1994). To be sure, it is generally expected that DEA and SFR results will differ as a consequence of the deterministic versus stochastic structure of the two approaches. Because hospital demand is subject to chance fluctuations, but hospital supply may respond only to peak demand, the influence of stochastic elements on efficiency measurement in this industry may actually be more pronounced than elsewhere in the economy. Likewise, it is generally recognized that DEA and SFR results are each sensitive to various underlying assumptions and the data used to operationalize them. For example, DEA findings may be sensitive to extreme data points, whereas SFR estimates

are expected to vary by the specific distributional assumption imposed on the composable error term. Clearly, the extent to which such matters actually confound the results of DEA and SFR hospital applications is highly uncertain at the moment. Comparative analyses are needed to narrow this uncertainty and to suggest agenda items for research that will advance the state-of-the-art.

The primary aim of this article is to compare the results of scoring hospital efficiency by means of DEA and SFR methods. A longitudinal data set on acute care hospital costs and service delivery in the state of Florida over the period 1982–1993 is assembled to carry out this task. A common set of output, input, and cost indicators is constructed; comparable DEA and SFR models are specified; and these models are estimated to obtain the efficiency values or scores yielded by each. We then assess whether the two models produce convergent or divergent evidence about hospital efficiency. In particular, we examine the degree to which the results of each model are similar in regard both to pegging the *industry* level of hospital inefficiency and in portraying the *individual* characteristics of the most and least efficient hospital facilities. We acknowledge at the outset that the analysis is neither extensive nor rigorous enough to document fully whether, and why, such DEA and SFR results should differ when applied to hospitals. Our more modest aim is to highlight some policy-relevant aspects of frontier methods for health services researchers and, thereby, to provide a point of departure for more detailed future work. Even at that, a substantial amount of technical detail has been suppressed in the text to afford a clearer view of those policy-related uses. For specialists interested in a more extensive account of our methodology, a technical appendix presents detailed descriptions of the data set, the choice and construction of the main variables, and the specifications of the empirical models; this appendix is available from the first author on request.

The article is organized as follows. The next section briefly summarizes the methods used to conduct the analysis. The subsequent section presents empirical estimates of the efficiency levels of Florida hospitals over a 12-year period. This section first sets out efficiency scores derived from estimating the basic DEA and SFR models; it also presents some correlates of DEA- and SFR-derived efficiency scores as a rough validity check on the results. Because substantial differences in SFR and DEA efficiency rankings are detected at this point of the analysis, additional empirical work is carried out. Included here are estimates of both different specifications of the SFR models and regression portraits testing whether DEA and SFR produce systematically different portraits of efficient and inefficient hospitals. The final section of the article sets out some policy implications of the analysis.

## METHODS

### *Data Set*

Since 1979, Florida has required that each hospital in the state submit various reports about financial performance on an annual basis for purposes of prospective budgetary review. Public use data tapes of these records covering the period 1982–1993 (inclusive) comprise the main source of data on service output indicators, inputs, operating expenses, and revenues for this study. In order to avoid the confounding effects of different types of hospital operations and industry turnover during the study period, we restrict the analysis to all short-term acute care Florida hospitals in continuous operation from 1982 through 1993. This produces a set of 186 hospitals for each year of the analysis and, correspondingly, an effective sample of 2,232 hospitals (or hospital-years) when these annual cross-sections are pooled. We refer to this pooled sample throughout as the set of *panel* hospitals.

### *Output and Input Variables*

Six output variables and a like number of input (cost) variables are constructed for each observation in the data set. The choice of these variables, as has long been the experience in the hospital literature, represents a compromise between what is ideal and what is feasible with the data at hand. In the case of outputs, we would like to use patient-level indicators of health outcomes to gauge the level and quality of final hospital production. Regrettably, suitably detailed information of this sort is unavailable for the entire 12-year study period; lack of data availability also explains why we exclude other types of final products of hospitals such as medical and allied health training outputs. In consequence, intermediate products that gauge the level and composition of patient care in the hospital essentially comprise our output vector ( $Q$ ). Even here, however, some compromises had to be made to derive feasible specifications of these intermediate output variables.

To begin with, indicators of inpatient output would ideally differentiate among service bundles that are admission-, stay-, and diagnosis-specific; given historical changes in reimbursement methodologies, we believe that they should also account for variations in the intensity of these service bundles stemming from different protocols and from economic inducements of major payer groups. We are able only to approximate these desiderata, and then only by keying inpatient output to the patient day. We distinguish the first day of care (admission day) from all others, supposing that differences in overall resource intensity attributable to diagnosis or the complexity/severity of cases

will be fixed on that day. The remaining days of care distinguish among major payer groups, whose reimbursement protocols may induce differing resource intensities of post-admission care, controlling for diagnosis. Operationally, then, we construct a case mix-weighted admissions variable (i.e., we scale total admissions in a given year by mean DRG weights for the same year) and three post-admission patient day variables (i.e., inpatient days net of the day of admission) corresponding to three payer categories: (1) Medicare; (2) Medicaid; and (3) Blue Cross, other private payers, and self-pay patients.

Ideally, the increasing importance and scope of hospital outpatient care would be represented in detail, distinguishing among others between services delivered to outpatients before or after an inpatient episode and services delivered to ambulatory patients who are not admitted to the hospital. As might be expected, the problem here is the large number of such outputs, all in different metrics. For this reason, we create two composite indexes of outpatient service activity. One is a composite index reflecting the provision of special tests and procedures (e.g., MRI, cardiac catheterization, physical therapy, etc.) to outpatients either before or after an inpatient episode; it is cast in admission-equivalent terms. The other measures the level of activity in ambulatory centers generating outpatient revenue in emergency room-equivalent terms.

Cost or annual expense figures ( $C$ ) are broken down by (1) wage and salary payments to personnel engaged in patient care activities (hospital service, ambulatory and ancillary activities); (2) wage and salary payments to personnel assigned to all nonpatient care centers (administration broadly defined); (3) other expenses in patient care cost centers; (4) capital costs (adjusted depreciation charges) for plant assets, that is, building and land; (5) adjusted depreciation charges for fixed and movable equipment; and (6) other nonpatient (administrative) costs attributable to capital use, made up of interest expense on long-term and short-term borrowings and all other expenses not elsewhere classified. These six cost categories are used directly in the DEA model; the sum of the six items is used to construct the dependent variable in the SFR model.

The vector of factor prices ( $W$ ) needed to implement the SFR model is constructed by dividing key inputs into corresponding annual expense categories. Three mean wage variables are obtained by dividing full-time equivalent (FTE) employment figures for inpatient/ambulatory care, ancillary patient care, and administrative personnel into their corresponding wage bills. The other three factor price variables gauge capital inputs: two divide annual depreciation charges by the corresponding book value of assets at

the beginning of each year for plant (buildings and land) and for fixed and movable equipment; the third divides total interest payments by the value of total current tangible and intangible assets yielding the implicit annual interest rate on debt financing instruments of the hospital.

Annual cost and factor price variables are scaled by a cross-sectional, state hospital price index that adjusts for nominal differences in input prices across local hospital markets. Because this article pools annual cross-sectional data, we further adjust this geographic index to reflect *intertemporal* changes in factor prices. We pieced together the PPS hospital input price index for the period 1982–1993 and used these figures to weight the annual means of the price index to approximate an intertemporal index.

### Model Specifications

The DEA model specification used in this analysis may be generally represented as:

$$\begin{aligned} \text{Max } \Theta_{n^*} &= \{ \Sigma_i \mu_i Q_{in^*} \} / \{ \Sigma_j v_j C_{jn^*} \} \\ \text{Subject to: } & \{ \{ \Sigma_i \mu_i Q_{in} \} / \{ \Sigma_j v_j C_{jn} \} \} \leq 1 \\ & (\mu_i, v_j \geq 0; i = 1, \dots, 6; j = 1, \dots, 6; n = 1, 2, \dots, N - 1). \end{aligned} \quad (1)$$

Here  $\Theta_{n^*}$  represents the weighted output/cost (input) ratio of reference hospital  $n^*$ ;  $Q_i$  and  $C_j$  represent the six output and six (input) cost variables described in the preceding section, all  $Q$  and  $C > 0$ ;  $n$  indexes sample hospitals exclusive of  $n^*$ , with  $N$  representing the total number of observations in any given sample partition; and  $\mu_i$  and  $v_j$  are the variable weights that maximize  $\Theta_{n^*}$  estimated by means of the fractional linear programming algorithm suggested by the work of Charnes, Cooper, and Rhodes (1978) and Boussofiane, Dyson, and Thanassoulis (1991). This linear programming formulation is fully described in the authors' technical appendix.

When Equation 1 is transformed into the linear programming algorithm and then estimated  $N$  times, treating each observation as  $n^*$  in turn, a  $\Theta$  value for each sample observation in the data set is obtained. Frontier-efficient hospitals receive a  $\Theta$  value of one, whilst those off of the frontier receive values proportional to the most efficient units, that is,  $0 \leq \Theta < 1$  for inefficient units, the degree of inefficiency characterized by ever lower values in this interval. Put differently, a given hospital with  $\Theta = 0.8$  effectively has costs that are 20 percent higher than it would have were it frontier-efficient, all else equal.

In order to facilitate the comparative analysis, the SFR regression model is specified to be as consistent with the DEA model as possible. We

rearrange the cost elements to obtain the conventional econometric total cost function,  $TC = F(Q_i, W_j)$ , where  $Q_i$  are the six outputs and  $W_j$  the six factor price variables defined above. In order to portray non-linearities in the cost consequences of changing output/input relationships as realistically as possible, we use the translog function because it portrays these non-linearities in a completely flexible way. A prototypical translog formulation applied to the SFR modeling strategy here is:

$$tc = f \left\{ q_i, w_j, \frac{1}{2} (q_i q_i, q_i q_k, w_j w_j, w_j w_m, q_j w_m); \beta, v + u \right\} \quad (2)$$

where the lowercase letters are the natural logarithms of the uppercase  $Q_i$  and  $W_j$  variables;  $i$  and  $k$  index the output variables ( $i \neq k$ );  $j$  and  $m$  index the factor price variables ( $j \neq m$ ); and  $\beta$  is the parameter vector and  $(v + u)$  the composable error to be estimated by maximum likelihood.

As Equation 2 suggests, the translog specification encompasses a large number of squared and cross-product terms—as a rule of thumb, about  $\{(q + w)(q + w + 1)/2\}$  number of such terms in an estimating equation with  $q$  number of output and  $w$  number of factor price variables. We began the SFR analysis by estimating a model with all squared and cross-product terms as in Equation 2, that is, a model with 91 parameters in all. (We refer below to this model as the *Full* translog.) For reasons detailed in the technical appendix, we then also estimated several more structured models as well, the results of two of which are reported below. One of these, labeled the *Basic* model, is a Cobb-Douglas hybrid that reduces the number of parameters to be estimated by imposing restrictions on the relationship between costs and factor prices and by creating a “common” set of higher-order terms for a select subgroup of outputs. The other specification was suggested when DEA and SFR efficiency scores were initially found to differ considerably. These differences appeared to stem from scale returns that vary parametrically in the SFR model but are implicitly treated as “constant” in our DEA specification. Accordingly, we specified a more structured version of the translog total cost function that constrains both output and factor price vectors to exhibit linear homogeneity or constant returns to scale in order to test the proposition that observed DEA and SFR differences arise from the way scale effects are treated. (We label this the *CRS* model.)

### *Efficiency Scoring*

As noted earlier, estimating the parameters of the SFR translog cost function is simply the first of two steps required to obtain the inefficiency residual. The second step uses the estimated regression parameters to compute mean

inefficiency at the level of the industry. The expected value of  $u$  for each individual observation is then calculated, conditional on the composable error and the assumption that the half-normal distribution governs the behavior of  $u$  (cf. Jondrow et al. 1982). Given the translog-type specification, residual inefficiency is computed as the proportional difference between the costs of a given hospital and the frontier cost level; correspondingly,  $u$  (actually, the antilog  $\exp u$ ) is scaled from zero upward. Recall that DEA efficiency is computed in mirror image terms; that is, most efficient hospitals take the value of one. In order to simplify the narrative from this point on, we invert the SFR residual so that it is nominally scaled in the same direction as the computed DEA value. We refer to the inverted residual as the frontier *score* and, when comparing it to the DEA value, refer to both as efficiency *scoring*.

### *Sample Subsets*

The data set on Florida hospitals is used in several different ways to facilitate the comparison between the DEA and SFR results. To begin with, we estimate the just described *Basic* SFR and DEA models cross-sectionally for each year of the 12-year period covered by the data. Then, because it is unclear that levels of efficiency can be effectively traced when the *frontier* or technological regime tapped in the estimation is permitted to change annually, we also pool the 12 annual cross-sections to obtain the longitudinal or panel set of 2,232 hospital-year observations. This pooling implicitly assumes that hospital services were produced under the same technological regime over the entire 12-year period. In order to examine the extent to which this assumption about technology influences the results, we compare annual cross-sectional results with panel estimates conditioned on individual years, that is, summary statistics computed for subsets of panel efficiency scores corresponding to individual years. Thus, for each model, our estimation yields 25 different vectors of efficiency scores, 12 for each annual cross-section ( $N = 186$ ), 12 for each conditioned panel year from the pooled data ( $N = 186$ ), and one from the pooled data set of the cross-sections ( $N = 2, 232$ ).

## EMPIRICAL FINDINGS

### *Efficiency Scores*

Table 1 sets out selected results from estimating the DEA and SFR models. The top panel presents estimated efficiency scores for a representative subset of years as well as for the 1982–1993 pooled data. These estimates suggest generally that Florida hospitals use resources inefficiently and that

this condition has not changed much over time. Yet it is immediately apparent that modeling and sample partitions influence this inference. Annual cross-sectional estimates, for example, tend generally to be higher than their panel counterparts, especially for the DEA model. This implies that assumptions about the underlying technology and the length of time for which it constrains output/input decisions make a difference to the results. Furthermore, variances of the efficiency distributions change between estimating models, as evidenced in Table 1 by coefficients of variation (C.V.) that differ by one order of magnitude for any given model and by several such orders across models.

Perhaps more interesting are the correlation coefficients presented in the bottom panel of Table 1. Observe first that the correlations between the cross-sectional DEA and SFR efficiency scores never exceed 0.4. (The coefficients

Table 1: Summary Statistics and Correlation Coefficients of Selected DEA and SFR Efficiency Scores (selected years, 1982–1993)

Statistics/Coefficients	Selected Cross-Sections						Pooled 1982–1993
	1982	1984	1986	1988	1990	1993	
Efficiency Scores* (Percent)							
DEA cross-section							
Mean	93.9	97.5	96.0	97.2	97.0	97.4	96.8
C.V.	10.1	6.2	9.5	5.9	7.0	6.9	7.5
DEA panel							
Mean	90.3	82.7	79.1	79.9	74.9	78.2	80.1
C.V.	12.6	14.8	15.4	17.5	20.6	21.6	18.5
SFR cross-section							
Mean	89.0	89.1	86.4	91.6	82.3	81.7	84.6
C.V.	5.7	5.7	9.3	2.9	15.1	17.9	14.8
SFR panel							
Mean	86.6	83.7	81.3	81.8	81.1	82.2	82.0
C.V.	10.2	11.5	12.7	13.8	14.7	13.3	14.4
Pearson Correlation Coefficients†							
DEA–SFR cross-sections	0.25	0.13	0.14	0.29	0.37	0.39	0.19
DEA–SFR panels	0.19	0.19	0.29	0.30	0.27	0.25	0.26
DEA cross-section–panel	0.70	0.52	0.37	0.53	0.38	0.34	0.36
SFR cross-section–panel	0.87	0.91	0.83	0.88	0.91	0.93	0.80
N	186	186	186	186	186	186	2232

\*See text for descriptions of the DEA model and the *Basic* version of the SFR model used to derive these efficiency scores as well as the cross-sectional and panel sample partitions. Like the efficiency means, the coefficients of variation (C.V.) are percentages.

†Unless otherwise noted, all correlation coefficients differ significantly from zero,  $p < .001$ .

for the annual cross-sections omitted in Table 1 are also all less than 0.4.) When the 12 yearly cross-sectional scores are pooled, the correlation coefficient is lower still. Weak correlations are also in evidence when the year-conditioned scores from the panel estimates are used in the computation; for example, the Pearson correlation between the DEA and SFR scores derived from the pooled model ( $N = 2,232$ ) conditioned on 1986 is only 0.29 and on 1993 only 0.25. Clearly, DEA and SFR efficiency scores do not map unto each other at all well.

Technological regime and sample size also appear to influence the efficiency scores derived from either type of frontier method, as evidenced by correlations between the cross-sectional and panel estimates of each model that are lower than might have been expected. The coefficient between the vector of DEA scores computed from the 1982 cross-section ( $N = 186$ ) and the scores for the larger pooled run ( $N = 2,232$ ) conditioned on 1982, for example, is only about 0.7, and this is the highest such correlation coefficient obtained, with the remainder of the years averaging only about 0.4. These intramodel correlations are higher for the SFR model, although the mapping between the cross-sectional and panel scores is hardly perfect. Furthermore, these SFR cross-sectional results do not have substantial predictive power. Only one cell in the full correlation matrix of all annual cross-sectional SFR scores (not reported here) exceeds 0.7, and only two others exceed 0.6.

We doubt that the observed differences between the SFR and DEA efficiency scores in Table 1 are due entirely to chance. This means that model choice and sample partition each influence the portrait of the efficient or inefficient hospital. As a means of investigating the extent to which they do, we next array the efficiency scores against a small set of policy-relevant hospital characteristics to assess how each model portrays the efficient and inefficient facility. (With the exception of the cost per case variable, which is adjusted for geographic variations in hospital input prices in any given year as well as an intertemporal deflator across each year of the 12-year period, this set of characteristics is commonplace and measured straightforwardly.) Because both models gauge inefficiency in relative terms, we establish arbitrary cut points on the distributions of the estimated scores in order to delineate subgroups of relatively more or less efficient facilities; to simplify the exposition, we focus on the distributions derived from the panel estimates of the DEA and SFR models for quartile cut points.

Table 2 shows the mean characteristics of hospitals whose DEA and SFR efficiency scores are in the top (highest) or bottom (lowest) quartiles of their respective distributions; as a point of comparison, it also shows the means of these characteristics across the entire panel sample of 2,232 hospital

observations. There are some similarities between the two models in regard to these efficiency subgroupings. Each, for instance, classifies hospitals with significantly lower real costs in its respective top quartile of scores; each also accords efficiency advantages to observations with shorter lengths of stay and lower FTE employment ratios. Somewhat unexpectedly, the two models identify government hospitals similarly as being *more* efficient than their for-profit and voluntary counterparts. Note in this regard that mean DRG weights differ across the efficiency distributions, with each model classifying hospitals with more severe case mixes in the most inefficient quartile, and vice versa.

Despite these similarities, several anomalies exist that perhaps suggest reasons why the findings yielded by each model differ. For one thing, the level of hospital activity indexed either by the number of cases (annual admissions) or the occupancy rate varies with the efficiency score in opposing directions. The SFR results suggest that smaller facilities and those with lower occupancy rates are more inefficient than those with more cases and higher occupancy; the DEA model suggests just the opposite. For another, the DEA model classifies facilities with larger bed complements as more inefficient than

Table 2: Selected Characteristics of Panel Hospitals Classified by DEA and SFR Efficiency Scores

<i>Characteristics</i>	<i>DEA Scores</i>		<i>SFR Scores<sup>†</sup></i>		<i>All Sample Hospitals</i>
	<i>Top Quartile</i>	<i>Bottom Quartile</i>	<i>Top Quartile</i>	<i>Bottom Quartile</i>	
Cases (mean annual number)	5987	10658*	8536	6989*	8287
Case mix (mean DRG weight) * 100	102	119*	104	113*	109
Control: government (%)	19	13*	22	11*	17
proprietary (%)	40	31*	44	41	41
voluntary (%)	41	56*	35	48*	42
Cost per case (mean \$)	3520	4662*	3167	4779*	3903
FTEs per 1000 cases (mean number)	86	106*	77	109*	91
Length of stay (mean days)	6	7*	6	7*	7
Licensed beds (mean number)	200	381*	261	270	282
Occupancy (mean %)	53	56*	58	52*	56
Teaching status (mean %)	9	21*	6	18*	12
<i>N</i>	558	558	558	558	2232

\*Significantly different from most efficient quartile,  $p \leq .001$ .

<sup>†</sup>The *Basic* version of the SFR model was used to estimate these scores; see text.

those with fewer licensed beds, and this difference is statistically significant; in contrast, the SFR model shows only a statistically insignificant relationship between bed size and efficiency.

The findings in regard to cases, bed size, and occupancy rates may stem from the fact that the *Basic* SFR model accounts more explicitly for scale factors in its specification than does the DEA model, which in effect assumes constant returns to scale. They may also stem more simply from the extent to which structure is imposed on the data by this SFR specification. As noted earlier (and as described in more detail in the technical appendix), several additional SFR specifications were estimated in order to test these conjectures. Table 3 presents selected findings from these additional estimations, including results from the completely specified translog model (*Full* model) and an even more structured Constant Returns to Scale specification (*CRS* model).

As can be seen, SFR efficiency scores change when either more or less structure is imposed on the data. In particular, when all interaction and cross-product terms are included in the *Full* SFR model, mean efficiency rises and the dispersion around that mean falls; in contrast, the *CRS* version of the SFR model reduces measured efficiency and dramatically increases the variance. Nonetheless, these new estimates do not substantially improve the correspondence between the DEA and SFR efficiency scores. The correlation coefficients between the DEA and SFR models, for instance, change only slightly. Interestingly, the correlations between and among the several versions of the SFR model change as much as those between the DEA and those SFR models. Note, for instance, that the efficiency scores derived from the unstructured and most structured versions of the SFR model are themselves correlated only at the level of 0.7.

More interesting still is that the rank-ordering of efficiency scores changes very little across these differing specifications. To see this, the right-

Table 3: Efficiency Scores Derived from DEA and Alternative SFR Model Specifications, Selected Findings ( $N = 2,232$  for all estimates)

Model/ Specification	Scores (%)		Pearson Correlation Matrix				DEA-SFR Scores Overlap (N)	
			DEA	SFR (B)	SFR (F)	SFR (C)	Top Quartile	Bottom Quartile
	Mean	C.V.						
DEA	80.1	18.4	1.00	—	—	—	—	—
SFR								
<i>Basic</i> model	82.0	14.4	0.26	1.00	—	—	227	217
<i>Full</i> model	85.1	10.6	0.33	0.83	1.00	—	214	236
<i>CRS</i> model	75.1	23.7	0.13	0.82	0.71	1.00	184	171

most columns of Table 3 show the number of observations classified in the top (bottom) quartile by *both* the DEA and the respective SFR models simultaneously; that is, the DEA and SFR scores of these hospitals overlap. This rearrangement of the data serves several purposes. It shows how the efficient or inefficient hospital is portrayed when, instead of choosing between the two models, the results of each are simultaneously combined to delineate efficiency subgroupings. In the second row, for instance, note that only about one in five (444) observations of the entire panel sample, and only about four in ten of the *quartile* groupings themselves ( $N = 558$ ), are characterized as either highly efficient or inefficient by the DEA and *Basic* SFR models. Similar results for the other two SFR models are also seen, even though the least structured model tends to classify more *inefficient* hospitals alike than do either of the more structured models. In all cases, however, the number of *non-overlapping* cases exceeds the number of *overlapping* ones. The DEA and SFR models, in other words, are more likely to classify differently than they are to classify alike.

### *Regression Analysis*

If systematic differences exist in the characteristics of hospitals that DEA and SFR models classify differently, then the portrait of the efficient and/or inefficient hospital will vary significantly by choice of approach. We estimate some regression models that test this proposition directly, while casting light indirectly on the factors that may account for the discordant DEA and SFR findings. In Table 4, we present the results of selected probit models testing both classification and continuous differences between the *Basic* SFR and DEA models. The dependent variables for the models reported in the first two columns are constructed by assigning a value of one if the SFR and DEA models classify the hospital in the same efficiency quartile (the top or bottom ones) or a value of zero if the two models classify the hospital differently. In the third column, the dependent variable takes the value of one if the SFR score exceeds the corresponding DEA score; zero otherwise. Although these models draw on different subsets of the sample of scores (889 hospitals classified in the top quartile by one or both of the models, 899 classified in the bottom quartile by either or both models, and the entire sample of 2,232 in the case of the score difference), they are expected to yield convergent results.

The regressor variables in these probit models are generally straightforward. Because wages, output mix, and case severity play roles in the specification of the efficiency models themselves, we do not introduce these

Table 4: Selected Probit Regression Estimates Comparing DEA and SFR Efficiency Scores

Explanatory Variables	Net Effects* (Absolute t-ratios)		
	Overlapping SFR and DEA Scores		Difference in Scores
	Top Quartile (=1)	Bottom Quartile (=1)	(SFR – DEA) > 0 (=1)
Beds (licensed number)	-0.0002 (2.21)	-0.0002 (0.23)	0.0006 (8.50)
Occupancy Rate (%)	0.0025 (2.30)	-0.0031 (2.80)	0.0014 (1.81)
Length of Stay (days)	-0.0235 (1.66)	0.0248 (2.15)	-0.0199 (2.09)
Control: religious (=1)	0.0133 (0.17)	-0.0612 (1.18)	0.2425 (4.94)
proprietary (=1)	-0.0142 (0.42)	-0.0525 (1.51)	0.0433 (1.74)
government (=1)	0.1466 (3.56)	-0.0044 (0.09)	-0.0173 (0.55)
Teaching facility (=1)	-0.0068 (0.11)	0.0914 (2.26)	-0.0867 (2.33)
DRG weight: above local average (=1)	0.0437 (0.79)	0.1455 (3.43)	-0.0766 (2.19)
below local average (=1)	0.0922 (2.05)	0.0405 (0.81)	-0.1204 (3.43)
Service mix: above local average (=1)	-0.0042 (0.07)	-0.0355 (0.75)	-0.0743 (1.92)
below local average (=1)	-0.0526 (1.07)	-0.0665 (1.36)	0.0160 (0.45)
Wage rates: above local average (=1)	0.0446 (1.21)	-0.0437 (1.14)	0.0797 (2.91)
below local average (=1)	-0.1038 (2.50)	0.0223 (0.65)	-0.0102 (0.39)
Constant	-0.1769 (2.30)	-0.2162 (2.51)	0.0088 (0.15)
Model chi-squared	43.99	34.26	186.31
Dependent variable: mean	0.2553	0.2414	0.627
s.d.	(0.44)	(0.43)	(0.484)
N	889	899	2232

Note: The *Basic* version of the SFR model was used to estimate the SFR scores; see text for descriptions of the efficiency scoring methods and the construction of the dependent variables.

\*Net effects are partial derivatives calculated as  $\beta_i f(z)$ , where  $\beta_i$  is the maximum likelihood probit coefficient for the  $i$ th regressor, and  $f(z)$  is the probability density function of the unit normal distribution evaluated at the means of the regressors.

continuously measured variables into the probit models directly. Rather, we create dummy variables indicating whether the hospital is above or below average in respect to these characteristics relative to other facilities in the local hospital market. In the first two columns, then, a significantly positive coefficient on any given regressor variable indicates that both the DEA and SFR models classify hospitals with that characteristic the same, while a significantly negative coefficient signals characteristics that are more likely to be classified in opposing ways. Coefficients that are statistically indistinguishable from zero suggest that the characteristic in question is neither more nor less likely to be treated similarly by the two models, all else equal. Roughly analogous inferences can be drawn from the signs and significance levels of the regressors in the third column: significantly positive (negative) coefficients imply that the SFR model is more (less) likely than the DEA model to find hospitals with those characteristics efficient, while insignificant coefficients suggest that the two models treat hospitals the same.

The results in Table 4 confirm generally that the choice of technique influences the portraits of the efficient and inefficient hospital; they also provide an additional basis for concluding that the differences between the DEA and SFR efficiency scores are not simply due either to chance alone or to differing assumptions about scale effects. To begin with, note that there are variations across most characteristics, with only one case in which all coefficients are insignificant: facilities that are below average in respect to service mix. The more common pattern is for one technique to favor the efficiency of a given characteristic over the other, but to do so either on the high or low end of the distribution of scores. For instance, the SFR model is significantly more likely to find efficiency differences related to bed size, although the classification rankings are more likely to be discordant at the high end (top quartile) of the distribution than at the lower end. In contrast, the DEA model is more likely to find length of stay related to efficiency but tends to classify only hospitals at the high end of the distribution discordantly; it agrees with the SFR model at the low end (bottom quartile). Somewhat similar findings in regard to teaching hospitals are found, where both models peg highly inefficient teaching facilities the same, although they differ generally and over the range of the most efficient quartile of scores.

## DISCUSSION

Two major inferences may now be drawn from the comparative analysis of the DEA and SFR efficiency models, one at the level of industry trends and

the other at the level where the most efficient or inefficient hospital is profiled. The empirical results suggest that Florida hospitals over the study period had costs that were substantially higher than the frontier level of costs; they also suggest that this inefficiency condition did not improve much over time. This result is quite striking, all the more so when one considers that the Florida hospital industry has a relatively higher proportion of for-profit facilities than other places, that it has been on the leading edge of the spread of managed care in both the public and private sectors during the 1980s and early 1990s, and that the hospitals in the study sample were selected because they had successfully survived this turbulent period. Although we cannot rule out the possibilities that the effects of efficiency occur only with substantial time lags and/or that efficiency has improved since 1993, the data adduced here that suggest that mean inefficiency may be on the order of 15 percent are consistent with the meager evidence now available, cf. Zuckerman, Hadley, and Iezzoni (1994). What is noteworthy from the perspective of the present study is that this conclusion emerges whether one relies on either the DEA or the SFR results alone.

But even though the DEA and SFR scores track efficiency similarly at the overall level of the industry, they map only roughly unto each other at the level of individual observations. Correlations across scores generally yielded midrange coefficients, many not so low as to be unreasonable but most not high enough, either, to obviate concern. Correlates of SFR and DEA scores also showed distinctly different patterns, with notable sign reversals between some hospital characteristics such as bed size and occupancy rates. Some differences in efficiency scores were anticipated, because the SFR model incorporates stochastic factors while the DEA model does not. Chance fluctuations due, say, to infectious disease outbreaks or atypical spikes in tourist flows, might well have affected measured output/input relationships in Florida hospitals. Yet significant differences are observed between the DEA and SFR rankings not only on the annual cross-sections, but also across time in the pooled cross-sectional estimates, where short-term random shocks should be expected to have less impact. Moreover, the significantly different pattern of correlations that we find between various hospital characteristics and DEA scores relative to SFR scores is not entirely consistent with the view that statistical noise alone accounts for the observed differences between the two techniques. The probit regression results reported in the previous section confirm this view. Accordingly, even when we acknowledge that chance factors play a role in both the locus of the efficiency frontier and the relative deviation of any observation from that frontier, we believe that the differences

between SFR and DEA stem from something more profound than just the fact that one is stochastic and the other is deterministic in nature.

Among other things, the differences between the techniques suggest that hospital policymakers should exercise extreme caution in proceeding to use frontier modeling immediately or extensively. Clearly, policymakers cannot be indifferent to the choice of technique, especially if they intend to employ frontier methods to identify the attributes or threshold characteristics of the most (or least) efficient hospital for rate-setting or selective contracting purposes. Our data strongly suggest that the portrait of the most or least efficient facility will vary substantially depending on whether DEA or SFR efficiency scores alone are used to prepare it. Policymakers, of course, may choose initially to experiment with one model or the other, perhaps basing their choice either on how well the findings compare to others reported in the literature or on the specific aims to be served by the analysis. DEA models may be more useful in smaller-scale studies designed to judge specific efficiency-improving interventions in given hospital markets, whereas SFR models may be better suited to industry-wide investigations of efficiency determinants and policy effectiveness. In either case, we caution against the widespread application of either SFR or DEA modeling until such time as the field better understands how, and why, they portray efficient and inefficient hospitals as differently as they do.

Frontier techniques are clearly promising enough for us to accord them high priority on the research agenda. More work along the lines pursued here is initially required to ascertain whether similarly divergent results are obtained for other samples and data sets. If so, additional efforts are then needed to refine and extend the DEA and SFR modeling in ways that help pinpoint reasons why divergent results are obtained. These efforts should focus primarily on how alternative output specifications influence the results. Clearly, efficiency scores may well differ when final outputs, especially health-related outcomes, can be incorporated in the empirical models. The sensitivity of the results to alternative specifications of intermediate outputs also needs more detailed testing than was possible in this article. Subsidiary analyses (not reported here) suggest, for instance, that both the level and the *pattern* of hospital outputs shape DEA efficiency scores relative to those yielded by the SFR model. This may stem partly from the way in which we split case- and stay-related dimensions of inpatient output, a specification that admits to straightforward interpretation in a regression framework, but it may give rise to some ambiguity in interpreting the DEA estimates.<sup>2</sup> The impact of patterns of intermediate outputs may also explain why the *CRS* variant of the

translog specification did not improve the concordance between the scores yielded by the two models. Yet the structure necessarily imposed on the SFR cost functions and the underlying distribution of the composable error may also be wrong. Additional comparisons of the two approaches using different measures and specifications would thus cast more light on whether the divergent results reported here are artifacts of our methodology and data or the result of other factors.

## ACKNOWLEDGMENTS

The authors acknowledge the extremely helpful comments of two anonymous referees on earlier drafts of this article.

## NOTES

1. Representative of the available DEA studies of U.S. hospitals are Grosskopf and Valdmanis (1987), Morey, Fine, Loree, et al. (1992), and Ozcan and Luke (1993). The SFR analysis is set out in Zuckerman, Hadley, and Iezzoni (1994). Banker, Conrad, and Strauss (1986) compare DEA and a more traditional translog regression model.
2. We owe this point to one of the anonymous referees who stressed that because DEA optimizes observations one at a time, output measures in DEA models must reflect true managerial desiderata and make sense in terms of their respective marginal rates of transformation. Although we believe that our stay-related variables (post-admission days) generally satisfy these criteria because they index different payer groups and reimbursement formulas, we acknowledge that detailed DEA results pertaining to one or a handful of hospitals may not always be immediately transparent in this regard. For the reasons adduced in Table 4 and discussed in more detail in the technical appendix, however, we doubt that this specification of the output variables is alone responsible for the observed difference in the DEA and SFR efficiency scores.

## REFERENCES

- Aigner, D., C. A. Knox Lovell, and P. Schmidt. 1977. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6 (1): 21–37.
- Banker, R. D., R. F. Conrad, and R. P. Strauss. 1986. "A Comparative Application of Data Envelopment Analysis and Translog Methods: An Illustrative Study of Hospital Production." *Management Science* 21 (1): 30–44.

- Batavia, A. I., R. J. Ozminkowski, G. Gaumer, and M. Gabay. 1993. "Lessons for States in Inpatient Rate Setting Under the Boren Amendment." *Health Care Financing Review* 15 (2): 137-54.
- Bauer, P. W. 1990. "Recent Developments in the Econometric Estimation of Frontiers." *Journal of Econometrics* 46 (1-2): 39-56.
- Boussofiane, A., R. G. Dyson, and E. Thanassoulis. 1991. "Applied Data Envelopment Analysis." *European Journal of Operational Research* 52 (1): 1-15.
- Charnes, A., W. W. Cooper, and E. Rhodes. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research* 2 (6): 429-44.
- Cowing, T. G., A. G. Holtmann, and S. Powers. 1983. "Hospital Cost Analysis: A Survey and Evaluation of Recent Studies." *Advances in Health Economics and Health Services Research* 4: 257-304.
- Greene, W. H. 1992. *LIMDEP, Version 6.0, User's Manual*. Bellport, NY: Econometric Software, Inc.
- Grosskopf, S., and V. Valdmanis. 1987. "Measuring Hospital Performance: A Non-parametric Approach." *Journal of Health Economics* 6 (2): 89-108.
- Hadley, J., and S. Zuckerman. 1994. "The Role of Efficiency Measurement in Hospital Rate Setting." *Journal of Health Economics* 13 (3): 335-40.
- Jondrow, J., C. A. Knox Lovell, I. S. Materov, and P. Schmidt. 1982. "On the Estimation of Technical Efficiency in the Stochastic Frontier Production Function Model." *Journal of Econometrics* 19 (2-3): 233-38.
- Kooreman, P. 1994. "Data Envelopment Analysis and Parametric Frontier Estimation: Complementary Tools." *Journal of Health Economics* 13 (3): 345-46.
- Morey, R. C., D. J. Fine, S. T. Loree, D. L. Retzlaff-Roberts, and S. Tsubakitani. 1992. "The Trade-Off Between Hospital Cost and Quality of Care: An Exploratory Empirical Analysis." *Medical Care* 30 (8): 677-98.
- Newhouse, J. P. 1994. "Frontier Estimation: How Useful a Tool for Health Economics?" *Journal of Health Economics* 13 (3): 317-22.
- Ozcan, Y., and R. Luke. 1993. "A National Study of the Efficiency of Hospitals in Urban Markets." *Health Services Research* 27 (6): 719-40.
- Zuckerman, S., J. Hadley, and L. Iezzoni. 1994. "Measuring Hospital Efficiency with Frontier Cost Functions." *Journal of Health Economics* 13 (3): 255-80.